Nonparametric estimation of the distribution of the autoregressive coefficient from panel random-coefficient AR(1) data

Remigijus Leipus, Anne Philippe, Vytautė Pilipauskaitė, Donatas Surgailis

▶ To cite this version:

HAL Id: hal-01174833
https://hal.archives-ouvertes.fr/hal-01174833

Submitted on 9 Jul 2015

Nonparametric estimation of the distribution of the autoregressive coefficient from panel random-coefficient AR(1) data[*]

Remigijus Leipus[1], Anne Philippe[2], Vytautė Pilipauskaitė[1,2], Donatas Surgailis[1]

July 9, 2015

[1]Vilnius University and [2]Université de Nantes

Abstract

We discuss nonparametric estimation of the distribution function $G(x)$ of the autoregressive coefficient from a panel of N random-coefficient AR(1) data, each of length n, by the empirical distribution of lag 1 sample correlations of individual AR(1) processes. Consistency and asymptotic normality of the empirical distribution function and a class of kernel density estimators is established under some regularity conditions on $G(x)$ as N and n increase to infinity. A simulation study for goodness-of-fit testing compares the finite-sample performance of our nonparametric estimator to the performance of its parametric analogue discussed in Beran *et al.* (2010).

Keywords: random-coefficient autoregression, empirical process, Kolmogorov-Smirnov statistic, kernel density estimator.

2010 MSC: 62G10, 62M10, 62G07.

1   Introduction

Panel data can describe a huge population of heterogeneous units/agents which evolve over the time, e.g. households, firms, industries, countries, stock market indices. In this paper we consider a panel where each individual unit evolves according to order-one random coefficient autoregressive model (RC AR(1)). It is well known that aggregation of specific RC AR(1) models can explain long memory phenomenon, which is often empirically observed in economic time series (see Granger (1980) for instance). More precisely, consider a panel $\{X_i(t), \, t = 1, \ldots, n, \, i = 1, \ldots, N\}$, where $X_i = \{X_i(t), t \in \mathbb{Z}\}$ is RC AR(1) process with the $(0, \sigma^2)$-noise and the random coefficient $a_i \in (-1, 1)$, whose autocovariance

$$\mathrm{E}X_i(0)X_i(t) \;\; = \;\; \sigma^2 \int_{-1}^{1} \frac{x^{|t|}}{1 - x^2} \, \mathrm{d}G(x) \tag{1.1}$$

is determined by the distribution $G(x) = \mathrm{P}(a \leq x)$ of the autoregressive coefficient. Granger (1980) showed, for a specific Beta-type distribution $G(x)$, that the *contemporaneous* aggregation of independent processes $\{X_i(t)\}$, $i = 1, \ldots, N$, results in a stationary Gaussian long memory process $\{\mathcal{X}(t)\}$, i.e.

$$N^{-1/2} \sum_{i=1}^{N} X_i(t) \; \to_{\mathrm{fdd}} \; \mathcal{X}(t) \quad \text{as } N \to \infty,$$

where the autocovariance $\mathrm{E}\mathcal{X}(0)\mathcal{X}(t) = \mathrm{E}X_1(0)X_1(t)$ decays slowly as $t \to \infty$ so that $\sum_{t \in \mathbb{Z}} |\mathrm{E}\mathcal{X}(0)\mathcal{X}(t)| = \infty$.

A natural statistical problem is recovering the correlation function in (1.1) or the distribution $G(x)$ (the frequency of a across the population of individual AR(1) 'microagents') from the aggregated sample $\{\mathcal{X}(t), t = 1, \ldots, n\}$. This problem was treated in Leipus *et al.* (2006), Celov *et al.* (2010), Chong (2006). Some related results were obtained in Celov *et al.* (2007), Horváth and Leipus (2009), and Jirak (2013). Albeit nonparametric, the estimators in Leipus *et al.* (2006) and Celov *et al.* (2010) involve an expansion of the density $g = G'$ in an orthogonal polynomial basis and are sensitive to the choice of the tuning parameter (the number of polynomials), being limited in practice to very smooth densities g. The last difficulty in estimation of G from aggregated data is not surprising due to the fact that aggregation *per se* inflicts a considerable loss of information about the evolution of individual 'micro-agents'.

Clearly, if the available data comprises evolutions $\{X_i(t), t = 1, \ldots, n\}, i = 1, \ldots, N$, of all N individual 'micro-agents' (the panel data), we may expect a much more accurate estimate of G. Robinson (1978) constructed an estimator for the moments of G using sample autocovariances of X_i and derived its asymptotic properties as $N \to \infty$, whereas the length n of each sample remains fixed. Beran *et al.* (2010) discussed estimation of two-parameter Beta densities g from panel AR(1) data using maximum likelihood estimators with unobservable a_i replaced by sample lag 1 autocorrelation of $X_i(1), \ldots, X_i(n)$ (see Sec. 5), and derived the asymptotic normality and some other properties of the estimators as N and n tend to infinity.

The present paper studies nonparametric estimation of G from panel random-coefficient AR(1) data using the empirical distribution function:

$$\widehat{G}_{N,n}(x) := \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(\widehat{a}_{i,n} \leq x), \quad x \in \mathbb{R}, \tag{1.2}$$

where $\widehat{a}_{i,n}$ is the lag 1 sample autocorrelation of $X_i, i = 1, \ldots, N$ (see (3.3)). We also discuss kernel estimation of the density $g(x) = G'(x)$ based on smoothed version of (1.2). We assume that individual AR(1) processes X_i are driven by identically distributed shocks containing both common and idiosyncratic (independent) components. Consistency and asymptotic normality as $N, n \to \infty$ of the above estimators are derived under some regularity conditions on $G(x)$. Our results can be applied to test goodness-of-fit of the distribution $G(x)$ to a given hypothesized distribution (e.g., a Beta distribution) using the Kolmogorov-Smirnov statistic, and to construct confidence intervals for $G(x)$ or $g(x)$.

The paper is organized as follows. Section 2 obtains the rate of convergence of the sample autocorrelation coefficient $\widehat{a}_{i,n}$ to a_i, in probability, the result of independent interest. Section 3 discusses the weak convergence of the empirical process in (1.2) to a generalized Brownian bridge. In Section 4 we study kernel density estimators of $g(x)$. We show that these estimates are asymptotically normally distributed and their mean integrated square error tends to zero. A simulation study of Section 5 compares the empirical performance of (1.2) and the parametric estimator of Beran *et al.* (2010) to the goodness-of-fit testing for $G(x)$ under null Beta distribution. The proofs of auxiliary statements can be found in the Appendix.

In what follows, C stands for a positive constant whose precise value is unimportant and which may change from line to line. We write $\to_{\mathrm{p}}, \to_{\mathrm{d}}, \to_{\mathrm{fdd}}$ for the convergence in probability and the convergence of (finite-dimensional) distributions respectively, whereas $\Rightarrow$ denotes the weak convergence in the space $D[-1, 1]$ with the supremum metric.

2 Estimation of random autoregressive coefficient

Consider a random-coefficient AR(1) process

$$X(t) \;=\; aX(t-1) + \zeta(t), \quad t \in \mathbb{Z}, \tag{2.1}$$

where innovations $\{\zeta(t)\}$ admit the following decomposition:

$$\zeta(t) \;=\; b\eta(t) + c\xi(t), \quad t \in \mathbb{Z}, \tag{2.2}$$

where random sequences $\{\eta(t)\}$, $\{\xi(t)\}$ and random coefficients a, b, c satisfy the following conditions:

Assumption A$_1$ $\{\eta(t)\}$ are independent identically distributed (i.i.d.) random variables (r.v.s) with $\mathrm{E}\eta(0) = 0$, $\mathrm{E}\eta^2(0) = 1$, $\mathrm{E}|\eta(0)|^{2p} < \infty$ for some $p > 1$.

Assumption A$_2$ $\{\xi(t)\}$ are i.i.d. r.v.s with $\mathrm{E}\xi(0) = 0$, $\mathrm{E}\xi^2(0) = 1$, $\mathrm{E}|\xi(0)|^{2p} < \infty$ for the same p as in A$_1$.

Assumption A$_3$ b and c are possibly dependent r.v.s such that $\mathrm{P}(b^2 + c^2 > 0) = 1$ and $\mathrm{E}b^2 < \infty$, $\mathrm{E}c^2 < \infty$.

Assumption A$_4$ $a \in (-1, 1)$ is a r.v. with a distribution function (d.f.) $G(x) := \mathrm{P}(a \leq x)$ supported on $[-1, 1]$ and satisfying

$$\mathrm{E}\Big[\frac{1}{1-|a|}\Big] \;=\; \int_{-1}^{1} \frac{\mathrm{d}G(x)}{1-|x|} \;<\; \infty. \tag{2.3}$$

Assumption A$_5$ a, $\{\eta(t)\}$, $\{\xi(t)\}$ and the vector $(b, c)'$ are mutually independent.

Remark 2.1 In the context of panel observations (see (3.1) below), $\{\eta(t)\}$ is the common component and $\{\xi(t)\}$ is the idiosyncratic component of shocks. The innovation process $\{\zeta(t)\}$ in (2.2) is i.i.d. if the coefficients b and c are nonrandom. In the general case $\{\zeta(t)\}$ is a dependent and uncorrelated stationary process with $\mathrm{E}\zeta(0) = 0$, $\mathrm{E}\zeta^2(0) = \mathrm{E}b^2 + \mathrm{E}c^2$, $\mathrm{E}\zeta(0)\zeta(t) = 0$, $t \neq 0$.

Under conditions A$_1$–A$_5$, a unique strictly stationary solution of (2.1) with finite variance exists and is written as

$$X(t) \;=\; \sum_{s \leq t} a^{t-s}\zeta(s), \quad t \in \mathbb{Z}. \tag{2.4}$$

Clearly, $\mathrm{E}X(t) = 0$ and $\mathrm{E}X^2(t) = \mathrm{E}\zeta^2(0)\mathrm{E}(1-a^2)^{-1} < \infty$. Note that (2.3) is equivalent to

$$\mathrm{E}\Big[\frac{1}{1-|a|^p}\Big] \;<\; \infty, \quad 1 < p \leq 2,$$

since $1 - |a| \leq 1 - |a|^p \leq 2(1-|a|)$ for $a \in (-1, 1)$.

For an observed sample $X(1), \ldots, X(n)$ from the stationary process in (2.4), define the sample lag 1 autocorrelation coefficient

$$\widehat{a}_n \;:=\; \frac{\sum_{t=1}^{n-1} X(t)X(t+1)}{\sum_{t=1}^{n} X^2(t)}. \tag{2.5}$$

Note $|\widehat{a}_n| \leq 1$ a.s. by the Cauchy inequality.

Proposition 2.1 *Under assumptions A_1–A_5, for any $0 < \gamma < 1$ and $n \geq 1$, it holds*

$$\mathrm{P}(|\widehat{a}_n - a| > \gamma) \;\leq\; C(n^{-(p/2)\wedge(p-1)}\gamma^{-p} + n^{-1}), \tag{2.6}$$

with $C > 0$ independent of n, γ.

PROOF. See Appendix.

Assume now that the d.f. $G(x) = \mathrm{P}(a \leq x)$ satisfies the following Hölder condition:

Assumption A_6 There exist constants $L_G > 0$ and $\varrho \in (0, 1]$ such that

$$|G(x) - G(y)| \;\leq\; L_G|x - y|^\varrho, \quad x, y \in [-1, 1]. \tag{2.7}$$

Consider the d.f. of $\widehat{a}_n$:

$$G_n(x) \;:=\; \mathrm{P}(\widehat{a}_n \leq x), \quad x \in \mathbb{R}. \tag{2.8}$$

Proposition 2.2 *Let assumptions A_1–A_6 hold. Then, for any $0 < \gamma < 1$ and $n \geq 1$,*

$$\sup_{x\in[-1,1]} |G_n(x) - G(x)| \;\leq\; L_G\gamma^\varrho + C(n^{-1} + n^{-(p/2)\wedge(p-1)}\gamma^{-p}) \tag{2.9}$$

with $C > 0$ independent of n, γ.

PROOF. Denote $\delta_n := \widehat{a}_n - a$. For any (nonrandom) $\gamma > 0$ we have

$$\sup_{x\in[-1,1]} |G_n(x) - G(x)| \;=\; \sup_{x\in[-1,1]} |\mathrm{P}(a + \delta_n \leq x) - \mathrm{P}(a \leq x)| \;\leq\; L_G\gamma^\varrho + \mathrm{P}(|\delta_n| > \gamma).$$

To see this, note that for any $\gamma > 0$ and *any* r.v. δ (possibly dependent on a)

$$\sup_{x\in\mathbb{R}} |\mathrm{P}(a + \delta \leq x) - \mathrm{P}(a \leq x)| \;\leq\; L_G\gamma^\varrho + \mathrm{P}(|\delta| > \gamma). \tag{2.10}$$

Indeed, using (2.7), which actually holds for all $x, y \in \mathbb{R}$, we have for any $x \in \mathbb{R}$ that

$$
\begin{aligned}
\mathrm{P}(a + \delta \leq x) - \mathrm{P}(a \leq x) \;&=\; \mathrm{P}(a + \delta \leq x, |\delta| \leq \gamma) - \mathrm{P}(a \leq x) + \mathrm{P}(a + \delta \leq x, |\delta| > \gamma) \\
&\leq\; \mathrm{P}(a \leq x + \gamma) - \mathrm{P}(a \leq x) + \mathrm{P}(|\delta| > \gamma) \\
&\leq\; L_G\gamma^\varrho + \mathrm{P}(|\delta| > \gamma)
\end{aligned}
$$

by (2.7), and similarly $\mathrm{P}(a \leq x) - \mathrm{P}(a + \delta \leq x) \leq L_G\gamma^\varrho + \mathrm{P}(|\delta| > \gamma)$, which proves (2.10). It remains to apply Proposition 2.1. $\qquad\square$

Corollary 2.3 *Let assumptions A_1–A_6 hold. Then, as $n \to \infty$,*

$$\sup_{x\in[-1,1]} |G_n(x) - G(x)| = O(n^{-\frac{\varrho}{\varrho+p}((p/2)\wedge(p-1))}).$$

PROOF follows from Proposition 2.2 by taking $\gamma = \gamma_n = o(1)$ such that $\gamma_n^\varrho \sim n^{-(p/2)\wedge(p-1)}\gamma_n^{-p}$ and noting that the exponent $\frac{\varrho}{\varrho+p}((p/2) \wedge (p-1)) < 1$. $\qquad\square$

4

3 Asymptotics of the empirical distribution function

Consider random-coefficient AR(1) processes $\{X_i(t)\}$, $i = 1, 2, \ldots$, which are stationary solutions to

$$X_i(t) = a_i X_i(t-1) + \zeta_i(t), \quad t \in \mathbb{Z}, \tag{3.1}$$

with innovations $\{\zeta_i(t)\}$ having the same structure as in (2.2):

$$\zeta_i(t) = b_i \eta(t) + c_i \xi_i(t), \quad t \in \mathbb{Z}. \tag{3.2}$$

More precisely, we make the following assumption:

Assumption B $\{\eta(t)\}$ satisfies A_1; $\{\xi_i(t)\}$, $(b_i, c_i)'$, a_i, $i = 1, 2, \ldots$, are independent copies of $\{\xi(t)\}$, $(b, c)'$, a which satisfy assumptions A_2–A_6. (Note that we assume A_5 for any $i = 1, 2, \ldots$.)

Define the corresponding sample correlation coefficients

$$\widehat{a}_{i,n} := \frac{\sum_{t=1}^{n-1} X_i(t) X_i(t+1)}{\sum_{t=1}^{n} X_i^2(t)} \tag{3.3}$$

and the empirical d.f.

$$\widehat{G}_{N,n}(x) := \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(\widehat{a}_{i,n} \le x), \quad x \in \mathbb{R}. \tag{3.4}$$

Recall that (3.4) is a nonparametric estimate of the d.f. $G(x) = \mathrm{P}(a_i \le x)$ from observed panel data $\{X_i(t), t = 1, \ldots, n, i = 1, \ldots, N\}$. In the following theorem we show that $\widehat{G}_{N,n}(x)$ is an asymptotically unbiased estimator of $G(x)$, as n and N both tend to infinity, and prove the weak convergence of the corresponding empirical process.

Theorem 3.1 *Assume the panel data model in* (3.1)–(3.2). *Let Assumption B hold and* $N, n \to \infty$. *Then*

$$\sup_{x \in [-1,1]} |\mathrm{E}\widehat{G}_{N,n}(x) - G(x)| = O(n^{-\frac{\varrho}{\varrho+p}((p/2)\wedge(p-1))}). \tag{3.5}$$

If, in addition, $N = o(n^{\frac{2\varrho}{\varrho+p}((p/2)\wedge(p-1))})$, *then*

$$N^{1/2}(\widehat{G}_{N,n}(x) - G(x)) \Rightarrow W(x), \quad x \in [-1, 1], \tag{3.6}$$

where $\{W(x), x \in [-1, 1]\}$ *is a continuous Gaussian process with zero mean and covariance* $\mathrm{Cov}(W(x), W(y)) = G(x \wedge y) - G(x)G(y)$, $x, y \in [-1, 1]$, *and* $W(-1) = W(1) = 0$.

PROOF. Note $\widehat{a}_{i,n}$, $i = 1, \ldots, N$, are identically distributed, in particular, $\mathrm{E}\widehat{G}_{N,n}(x) = G_n(x)$ with $G_n(x)$ defined in (2.8). Hence, (3.5) follows immediately from Corollary 2.3.

To prove the second statement of the theorem, we approximate $\widehat{G}_{N,n}(x)$ by the empirical d.f.

$$\widehat{G}_N(x) := \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(a_i \le x), \quad x \in [-1, 1]$$

of i.i.d. r.v.s $a_i, i = 1, \ldots, N$. We have $N^{1/2}(\widehat{G}_{N,n}(x) - G(x)) = N^{1/2}(\widehat{G}_N(x) - G(x)) + D_{N,n}(x)$ with $D_{N,n}(x) := N^{1/2}(\widehat{G}_{N,n}(x) - \widehat{G}_N(x))$. Since A_6 guarantees the continuity of G, it holds

$$N^{1/2}(\widehat{G}_N(x) - G(x)) \Rightarrow W(x), \quad x \in [-1, 1]$$

by the classical Donsker's theorem, and (3.6) follows from $\sup_{x\in[-1,1]}|D_{N,n}(x)| \to_p 0$. By definition,

$$D_{N,n}(x) = N^{-1/2}\sum_{i=1}^{N}(\mathbf{1}(a_i + \delta_{i,n} \le x) - \mathbf{1}(a_i \le x)) = D'_{N,n}(x) - D''_{N,n}(x),$$

where $\delta_{i,n} := \widehat{a}_{i,n} - a_i$, $i = 1, \ldots, N$, and

$$D'_{N,n}(x) := N^{-1/2}\sum_{i=1}^{N}\mathbf{1}(x < a_i \le x - \delta_{i,n}, \delta_{i,n} \le 0),$$

$$D''_{N,n}(x) := N^{-1/2}\sum_{i=1}^{N}\mathbf{1}(x - \delta_{i,n} < a_i \le x, \delta_{i,n} > 0).$$

For $\gamma > 0$ we have

$$D'_{N,n}(x) \le N^{-1/2}\sum_{i=1}^{N}\mathbf{1}(x < a_i \le x + \gamma) + N^{-1/2}\sum_{i=1}^{N}\mathbf{1}(|\delta_{i,n}| > \gamma) =: V'_N(x) + V''_{N,n}.$$

(Note that $V''_{N,n}$ does not depend on x.) By Proposition 2.1, we obtain

$$\mathrm{E}V''_{N,n} = N^{-1/2}\sum_{i=1}^{N}\mathrm{P}(|\delta_{i,n}| > \gamma) \le CN^{1/2}(n^{-(p/2)\wedge(p-1)}\gamma^{-p} + n^{-1}),$$

which tends to 0 when γ is chosen as $\gamma^{\varrho+p} = n^{-(p/2)\wedge(p-1)} \to 0$. Next,

$$V'_N(x) = N^{1/2}(\widehat{G}_N(x+\gamma) - \widehat{G}_N(x)) = N^{1/2}(G(x+\gamma) - G(x)) + U_N(x, x+\gamma],$$

$$U_N(x, x+\gamma] := N^{1/2}(\widehat{G}_N(x+\gamma) - G(x+\gamma)) - N^{1/2}(\widehat{G}_N(x) - G(x)).$$

The above choice of $\gamma^{\varrho+p} = n^{-(p/2)\wedge(p-1)}$ implies $\sup_{x\in[-1,1]} N^{1/2}|G(x+\gamma) - G(x)| = O(N^{1/2}\gamma^{\varrho}) = o(1)$, whereas $U_N(x, x+\gamma]$ vanishes in the uniform metric in probability (see Lemma 6.2 in Appendix). Since $D''_{N,n}(x)$ is analogous to $D'_{N,n}(x)$, this proves the theorem. $\qquad\square$

Remark 3.1 Theorem 3.1 can be used for testing goodness-of-fit, i.e. the null hypothesis $H_0\colon G = G_0$ vs. $H_1\colon G \neq G_0$ with G_0 being a certain hypothetical distribution satisfying the Hölder condition in (2.7). Accordingly, the corresponding Kolmogorov-Smirnov test rejecting H_0 whenever

$$N^{1/2}\sup_{x\in[-1,1]}|\widehat{G}_{N,n}(x) - G_0(x)| > c_\alpha \tag{3.7}$$

has asymptotic size $\alpha \in (0,1)$ provided N, n, G_0 satisfy the assumptions for (3.6) in Theorem 3.1. (Here, c_α is the upper α-quantile of the Kolmogorov distribution.) See also Section 5. One can also consider other statistics that are continuous functionals of the empirical process $\{N^{1/2}(\widehat{G}_{N,n}(x) - G_0(x)), x \in [1,1]\}$.

4 Kernel density estimation

In this section we assume G has a bounded probability density function $g(x) = G'(x), x \in [-1,1]$, implying assumption A$_6$ with Hölder exponent $\varrho = 1$ in (2.7). It is of our interest to estimate $g(x)$ in a nonparametric way from $\widehat{a}_{1,n}, \ldots, \widehat{a}_{N,n}$ (3.3).

Consider the kernel density estimator

$$\widehat{g}_{N,n}(x) \; := \; \frac{1}{Nh}\sum_{i=1}^{N}K\Big(\frac{x-\widehat{a}_{i,n}}{h}\Big), \quad x \in \mathbb{R}, \tag{4.1}$$

where K is a kernel, satisfying Assumption A_7 and $h = h_{N,n}$ is a bandwidth which tends to zero as N and n tend to infinity.

Assumption A_7 $\; K : [-1,1] \to \mathbb{R}$ is a continuous function of bounded variation that satisfies $\int_{-1}^{1}K(x)\mathrm{d}x = 1$. Set $\|K\|_2^2 := \int_{-1}^{1}K(y)^2\mathrm{d}y$ and $\mu_2(K) := \int_{-1}^{1}y^2K(y)\mathrm{d}y$ and $K(x) := 0$, $x \in \mathbb{R} \setminus [-1,1]$.

We consider two cases separately.

Case (i) $\mathrm{P}(b_1 = 0) = 1$, meaning that the coefficient $b_i = 0$ for the common shock in (3.2) is zero and that the individual processes $\{X_i(t)\}$, $i = 1, 2, \ldots$, are independent and satisfy

$$X_i(t) \; = \; a_iX_i(t-1) + c_i\xi_i(t), \quad t \in \mathbb{Z}.$$

Case (ii) $\mathrm{P}(b_1 \neq 0) > 0$, meaning that $\{X_i(t)\}$, $i = 1, 2, \ldots$, are mutually dependent processes.

Proposition 4.1 *Let assumptions B and A_7 hold. If $h^{1+p}n^{(p/2)\wedge(p-1)} \to \infty$, then*

$$\mathrm{E}\widehat{g}_{N,n}(x) \; \to \; g(x) \tag{4.2}$$

at every continuity point $x \in \mathbb{R}$ of g. Moreover, if

$$\begin{cases} n^{(p/2)\wedge(p-1)}h^{1+p} \to \infty & \text{in Case (i),} \\ n^{(p/2)\wedge(p-1)}(h/N)^{1+p} \to \infty & \text{in Case (ii),} \end{cases} \tag{4.3}$$

then

$$Nh\mathrm{Cov}(\widehat{g}_{N,n}(x_1), \widehat{g}_{N,n}(x_2)) \; \to \; \begin{cases} g(x_1)\|K\|_2^2 & \text{if } x_1 = x_2, \\ 0 & \text{if } x_1 \neq x_2 \end{cases} \tag{4.4}$$

at any continuity points $x_1, x_2 \in \mathbb{R}$ of g. If $Nh \to \infty$ holds in addition to (4.3), then the estimator $\widehat{g}_{N,n}(x)$ is consistent at each continuity point $x \in \mathbb{R}$:

$$\mathrm{E}|\widehat{g}_{N,n}(x) - g(x)|^2 \; \to \; 0. \tag{4.5}$$

PROOF. Throughout the proof, let $K_h(x) := K(x/h)$, $x \in \mathbb{R}$. Consider (4.2). Note $\mathrm{E}\widehat{g}_{N,n}(x) = h^{-1}\mathrm{E}K_h(x - \widehat{a}_n)$, because $\widehat{a}_{i,n}$, $i = 1, \ldots, N$, are identically distributed. Let us approximate $\widehat{g}_{N,n}(x)$ by

$$\widehat{g}_N(x) \; := \; \frac{1}{Nh}\sum_{i=1}^{N}K_h(x - a_i), \quad x \in \mathbb{R}, \tag{4.6}$$

which satisfies $\mathrm{E}\widehat{g}_N(x) = h^{-1}\mathrm{E}K_h(x - a) \to g(x)$ as $h \to 0$ at a continuity point x of g, see Parzen (1962). Integration by parts and Corollary 2.3 yield

$$\begin{aligned} h|\mathrm{E}\widehat{g}_{N,n}(x) - \mathrm{E}\widehat{g}_N(x)| &= \Big|\int_{\mathbb{R}}(G_n(y) - G(y))\mathrm{d}K_h(x - y)\Big| \tag{4.7} \\ &\leq V(K)\sup_{z \in [-1,1]}|G_n(y) - G(y)| \; = \; O(n^{-((p/2)\wedge(p-1))/(1+p)}), \end{aligned}$$

uniformly in $x \in \mathbb{R}$, where $V(K)$ denotes the total variation of K and $V(K) = V(K_h)$. This proves (4.2).

Next, let us prove (4.4). We have

$$
Nh\mathrm{Cov}(\widehat{g}_N(x_1), \widehat{g}_N(x_2)) \;=\; \frac{1}{h}\, \mathrm{E}K_h(x_1 - a)K_h(x_2 - a) \;\to\;
\begin{cases}
g(x_1)\|K\|_2^2 & \text{if } x_1 = x_2, \\
0 & \text{if } x_1 \neq x_2,
\end{cases}
\tag{4.8}
$$

as $h \to 0$ at any points x_1, x_2 of continuity of g, see Parzen (1962). Split $Nh\{\mathrm{Cov}(\widehat{g}_{N,n}(x_1), \widehat{g}_{N,n}(x_2)) - \mathrm{Cov}(\widehat{g}_N(x_1), \widehat{g}_N(x_2))\} = \sum_{i=1}^{3} Q_i(x_1, x_2)$, where

$$
\begin{aligned}
Q_1(x_1, x_2) &:= h^{-1}[\mathrm{E}K_h(x_1 - \widehat{a}_n)K_h(x_2 - \widehat{a}_n) - \mathrm{E}K_h(x_1 - a)K_h(x_2 - a)], \\
Q_2(x_1, x_2) &:= h^{-1}[\mathrm{E}K_h(x_1 - \widehat{a}_n)\mathrm{E}K_h(x_2 - \widehat{a}_n) - \mathrm{E}K_h(x_1 - a)\mathrm{E}K_h(x_2 - a)], \\
Q_3(x_1, x_2) &:= (N-1)h^{-1}\,\mathrm{Cov}(K_h(x_1 - \widehat{a}_{1,n}), K_h(x_2 - \widehat{a}_{2,n})).
\end{aligned}
$$

Note $Q_3(x_1, x_2) = 0$ in Case (i). Similarly to (4.7), $|Q_1(x_1, x_2)| = h^{-1}|\int_{\mathbb{R}} (G_n(y) - G(y))\mathrm{d}K_h(x_1 - y)K_h(x_2 - y)| \leq Ch^{-1}n^{-((p/2)\wedge(p-1))/(1+p)} \to 0$ since $V(K_h(x_1 - \cdot)K_h(x_2 - \cdot)) \leq C$ and $|Q_2(x_1, x_2)| \leq Ch^{-1}n^{-((p/2)\wedge(p-1))/(1+p)} \to 0$ uniformly in x_1, x_2. Finally,

$$
\begin{aligned}
|Q_3(x_1, x_2)| &= (N-1)h^{-1}\left| \int_{\mathbb{R}} \int_{\mathbb{R}} (\mathrm{P}(\widehat{a}_{1,n} \leq y_1, \widehat{a}_{2,n} \leq y_2) - \mathrm{P}(\widehat{a}_{1,n} \leq y_1)\mathrm{P}(\widehat{a}_{2,n} \leq y_2))\mathrm{d}K_h(x_1 - y_1)\mathrm{d}K_h(x_2 - y_2) \right| \\
&\leq CNh^{-1} \sup_{y_1, y_2 \in [-1,1]} |\mathrm{P}(\widehat{a}_{1,n} \leq y_1, \widehat{a}_{2,n} \leq y_2) - \mathrm{P}(\widehat{a}_{1,n} \leq y_1)\mathrm{P}(\widehat{a}_{2,n} \leq y_2)| \\
&= O(Nh^{-1}n^{-((p/2)\wedge(p-1))/(1+p)}) = o(1),
\end{aligned}
$$

proving (4.4) and the proposition. $\qquad\square$

Remark 4.1 It follows from the proof of the above proposition that in the case of a (uniformly) continuous density $g(x), x \in [-1, 1]$ relations (4.2), (4.5) and the first relation in (4.4) hold uniformly in $x \in \mathbb{R}$, implying the convergence of the mean integrated squared error:

$$
\int_{-\infty}^{\infty} \mathrm{E}|\widehat{g}_{N,n}(x) - g(x)|^2 \, \mathrm{d}x \;\to\; 0.
$$

Proposition 4.2 (Asymptotic normality) *Let assumptions B and A_7 hold and assume $Nh \to \infty$ in addition to (4.3). Moreover, let K be a Lipschitz function in Case* (ii). *Then*

$$
\frac{\widehat{g}_{N,n}(x) - \mathrm{E}\widehat{g}_{N,n}(x)}{\sqrt{\mathrm{Var}(\widehat{g}_{N,n}(x))}} \;\to_{\mathrm{d}}\; \mathrm{N}(0, 1),
\tag{4.9}
$$

at every continuity point $x \in (-1, 1)$ of g.

PROOF. First, consider Case (i). Since $\widehat{g}_{N,n}(x) = (Nh)^{-1}\sum_{i=1}^{N} V_{i,N}$ is a (normalized) sum of i.i.d. r.v.s $V_{i,N} := K_h(x - \widehat{a}_{i,n})$ with common distribution $V_N := V_{1,N}$, it suffices to verify Lyapunov's condition

$$
\frac{\mathrm{E}|V_N - \mathrm{E}V_N|^{2+\delta}}{N^{\delta/2}\,[\mathrm{Var}(V_N)]^{(2+\delta)/2}} \;\to\; 0,
\tag{4.10}
$$

for some $\delta > 0$. This follows by the same arguments as in Parzen (1962). Analogously to Proposition 4.1, we have $\mathrm{E}|V_N|^{2+\delta} = \mathrm{E}|K_h(x - \widehat{a}_n)|^{2+\delta} \sim hg(x)\int_{-1}^{1}|K(y)|^{2+\delta}\mathrm{d}y = O(h)$ while $\mathrm{Var}(V_N) = Nh^2\mathrm{Var}(\widehat{g}_{N,n}(x)) \sim hg(x)\|K\|_2^2$ according to (4.4). Hence the l.h.s. of (4.10) is $O((Nh)^{-\delta/2}) = o(1)$, proving (4.9) in Case (i).

8

Let us turn to Case (ii). It suffices to prove that $\sqrt{Nh}(\widehat{g}_{N,n}(x) - \widehat{g}_N(x)) \to_p 0$, for $\widehat{g}_N(x)$ given in (4.6). By $|K(x) - K(y)| \le L_K|x - y|$, $x, y \in \mathbb{R}$, for $\epsilon > 0$

$$
\begin{aligned}
\mathrm{P}\Big(\sqrt{Nh}|\widehat{g}_{N,n}(x) - \widehat{g}_N(x)| > \epsilon\Big) &\le \mathrm{P}\Big(\frac{L_K}{\sqrt{Nh}}\sum_{i=1}^{N}\frac{|\widehat{a}_{i,n} - a_i|}{h} > \epsilon\Big) \\
&\le N\mathrm{P}\Big(|\widehat{a}_n - a| > \sqrt{Nh}\Big(\frac{h}{N}\Big)\frac{\epsilon}{L_K}\Big), \\
&\le C(h(Nh)^{-p/2}(N/h)^{1+p}n^{-(p/2)\wedge(p-1)} + (N/n)) = o(1)
\end{aligned}
$$

follows from Proposition 2.1 and (4.3) with $Nh \to \infty$. $\qquad\square$

Corollary 4.3 *Let assumptions of Proposition 4.2 hold with $h \sim cN^{-1/5}$ for some $c > 0$, i.e.*

$$
N = \begin{cases} o(n^{\frac{5}{3}\frac{1}{1+p}(\frac{p}{2}\wedge(p-1))}) & \text{in Case (i),} \\ o(n^{\frac{5}{6}\frac{1}{1+p}(\frac{p}{2}\wedge(p-1))}) & \text{in Case (ii).} \end{cases}
$$

Moreover, let $g \in C^2[-1,1]$ and $\int_{-1}^{1} yK(y)\mathrm{d}y = 0$. Then

$$
N^{2/5}(\widehat{g}_{N,n}(x) - g(x)) \to_d \mathrm{N}(\mu(x), \sigma^2(x)),
$$

where $\mu(x) := (c^2/2)g''(x)\mu_2(K)$ and $\sigma^2(x) := (1/c)g(x)\|K\|_2^2$.

PROOF follows from Proposition 4.2, by noting that $\mathrm{E}\widehat{g}_N(x) - g(x) \sim h^2 g''(x)\mu_2(K)/2$ as $h \to 0$ and $\mathrm{E}\widehat{g}_{N,n}(x) - \mathrm{E}\widehat{g}_N(x) = O(h^{-1}n^{-((p/2)\wedge(p-1))/(1+p)})$ by (4.7). $\qquad\square$

5 Simulations for goodnesss-of-fit testing

In this section we compare our nonparametric goodness-of-fit test in (3.7) for testing the null hypothesis $G = G_0$ with its parametric analogue studied in Beran *et al.* (2010). In accordance with the last paper, we assume $\{X_i(t)\}$ in (3.1) to be independent AR(1) processes with standard normal i.i.d. innovations $\{\zeta_i(t)\}$, $\zeta(0) \sim \mathrm{N}(0,1)$ and the random autoregressive coefficient $a_i \in (0,1)$ having a Beta-type density $g(x)$ with unknown parameters $\theta := (\alpha, \beta)'$:

$$
g(x) = \frac{2}{\mathrm{B}(\alpha,\beta)}x^{2\alpha-1}(1-x^2)^{\beta-1}, \quad x \in (0,1), \quad \alpha > 1, \beta > 1, \tag{5.1}
$$

where $\mathrm{B}(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ is Beta function. Note that $\beta \in (1,2)$ implies the long memory property in $\{X_i(t)\}$. Beran *et al.* (2010) discuss a maximum likelihood estimate $\widehat{\theta}_{N,n,\kappa} = (\widehat{\alpha}, \widehat{\beta})'$ of $\theta = (\alpha, \beta)'$ when each unobservable coefficient a_i is replaced by its estimate $\widehat{a}_{i,n,\kappa} := \min\{\max\{\widehat{a}_{i,n}, \kappa\}, 1 - \kappa\}$ with $\widehat{a}_{i,n}$ given in (3.3) and $0 < \kappa = \kappa(N,n) \to 0$ is a truncation parameter. Under certain conditions on $N, n \to \infty$ and $\kappa \to 0$, Beran *et al.* (2010, Theorem 2) showed that

$$
N^{1/2}(\widehat{\theta}_{N,n,\kappa} - \theta_0) \to_d \mathrm{N}(0, A^{-1}(\theta_0)), \tag{5.2}
$$

where θ_0 is the true parameter vector,

$$
A(\theta) := \begin{pmatrix} \psi_1(\alpha) - \psi_1(\alpha+\beta) & -\psi_1(\alpha+\beta) \\ -\psi_1(\alpha+\beta) & \psi_1(\beta) - \psi_1(\alpha+\beta) \end{pmatrix},
$$

and $\psi_1(x) := \mathrm{d}^2 \ln \Gamma(x)/\mathrm{d}x^2$ is the Trigamma function. Based on (3.7) and (5.2), we consider testing both ways (nonparametrically and parametrically) the hypothesis that the unobserved autoregressive coefficients $\{a_1, \ldots, a_N\}$ are drawn from the reference distribution G_0 having density function in (5.1) with a specific θ_0, i.e. the null $G = G_0$ vs. the alternative $G \neq G_0$. The respective test statistics are

$$T_1 := N^{1/2} \sup_x |\widehat{G}_{N,n}(x) - G_0(x)| \quad \text{and} \quad T_2 := N(\widehat{\theta}_{N,n,\kappa} - \theta_0)' A(\theta_0)(\widehat{\theta}_{N,n,\kappa} - \theta_0). \tag{5.3}$$

Under the null hypothesis, the statistics T_1 and T_2 converge to the Kolmogorov distribution and the chi-square distribution with 2 degrees of freedom, respectively, see (3.7), (5.2).

To compare the performance of the above testing procedures, we compute the empirical distribution of the p-value of T_1 and T_2 under null and alternative hypotheses. The p-value of observed T_i is defined as $p(T_i) = 1 - \mathcal{K}_i(T_i), i = 1, 2$, where $\mathcal{K}_i(x), i = 1, 2$ denote the limit distribution functions of (5.3). Recall that when the significance level of the test is correct, the (asymptotic) distribution of the p-value is uniform on $[0, 1]$. The simulation procedure to compare the performance of T_1 and T_2 is the following:

Step S_0 We fix the parameter under the null hypothesis $H_0 : \theta = \theta_0$ with $\theta_0 = (2, 1.4)'$.

Step S_1 We simulate 5000 panels with $N = 250$, $n = 817$ for five chosen values $\theta = (2, 1.2)', (2, 1.3)',$ $(2, 1.4)', (2, 1.5)', (2, 1.6)'$ of Beta parameters.

Step S_2 For each simulated panel we compute the p-value of statistics T_1 and T_2.

Step S_3 The empirical c.d.f.'s of computed p-values of statistics T_1 and T_2 are graphed.

The values of Beta parameters $\theta_0 = (2, 1.4)'$, N, n were chosen in accordance with the simulation study in Beran *et al.* (2010).

Fig. 1 presents the simulation results under the true hypothesis $\theta = \theta_0$ with zoom-in on small p-values. We see that both c.d.f.'s in the left graph are approximately linear. Somewhat surprisingly, it appears that the empirical size of T_1 (the nonparametric test) is better than the size of T_2 (the parametric test). Particularly, for significance levels 0.05 and 0.1 we provide the empirical size values in Table 1.

Fig. 2 gives the graphs of the empirical c.d.f.'s of p-values of T_1 and T_2 for several alternatives $\theta \neq \theta_0$. It appears that for $\beta > \beta_0 = 1.4$ the parametric test T_2 is more powerful than the nonparametric test T_1 but for $\beta < \beta_0$ the power differences are less significant. Table 1 illustrates the empirical power for the significance levels 0.05, 0.1.

Signif. level	5%					10%				
β	1.2	1.3	1.4	1.5	1.6	1.2	1.3	1.4	1.5	1.6
T_1	.532	.137	.049	.208	.576	.653	.223	.103	.315	.702
T_2	.500	.104	.077	.313	.735	.634	.184	.134	.421	.827

Table 1: Numerical results of the comparison for testing procedure $H_0 : \theta = (2, 1.4)'$ at significance level 5% and 10% . The column for $\beta = 1.4$ provides the empirical size.

The above simulations (Fig. 1 and 2, Table 1) refer to the case of independent individual processes $\{X_i(t)\}$. There are no theoretical results for the parametric test T_2, when AR(1) series are dependent. Although the nonparametric test T_1 is valid for the latter case, one may expect that the presence of the common shock component in the panel data in (3.2) has a negative effect on the test performance for short series. To illustrate this effect, we simulate 5000 panels with AR(1) processes $\{X_i(t)\}$ driven by dependent shocks in
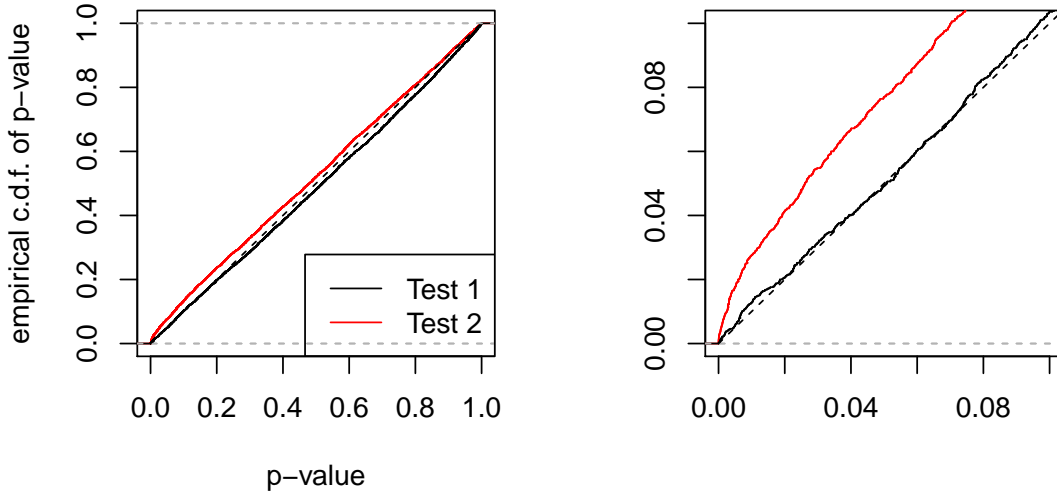
Figure 1: [left] Empirical c.d.f. of p-values of T_1 and T_2 under $H_0 : \theta_0 = (2, 1.4)'$; 5000 replications with $N = 250$, $n = 817$. [right] Zoom-in on the region of interest: p-values smaller than 0.1.

(3.2) with $b_i = b$, $c_i = (1 - b^2)^{1/2}$. As previously, we choose $\theta_0 = (2, 1.4)'$, $N = 250$, $n = 817$ and we fix $\theta = (2, 1.4)'$ to evaluate the empirical size of T_1. Fig. 3 [left] presents the graphs of the empirical c.d.f.'s of the p-values of T_1 for $b = 1$, $b = 0.6$ and $b = 0$, the latter corresponding to independent individual processes as in Fig. 1. We see that the size of the test worsens when b increases, particularly when $b = 1$ and the individual processes are all driven by the same common noise. To overcome the last effect, the sample length n of each series in the panel may be increased as in Fig. 3 [right], where the choice of $n = 5500$ and $b = 1$ shows a much better performance of T_1 under the null hypothesis $\theta = \theta_0 = (2, 1.4)'$ and the alternative ($\theta = (2, 1.5)'$ and $\theta = (2, 1.6)'$) scenarios.

In conclusion,

1. We do not observe an important loss of the power for the nonparametric KS test T_1 compared to the parametric approach.

2. The KS test T_1 does not require to choose any tuning parameter contrary to the test T_2.

3. One can use the KS test T_1 under weaker assumptions on AR(1) innovations. We only impose moment conditions. The dependence between the series is allowed by (3.2).
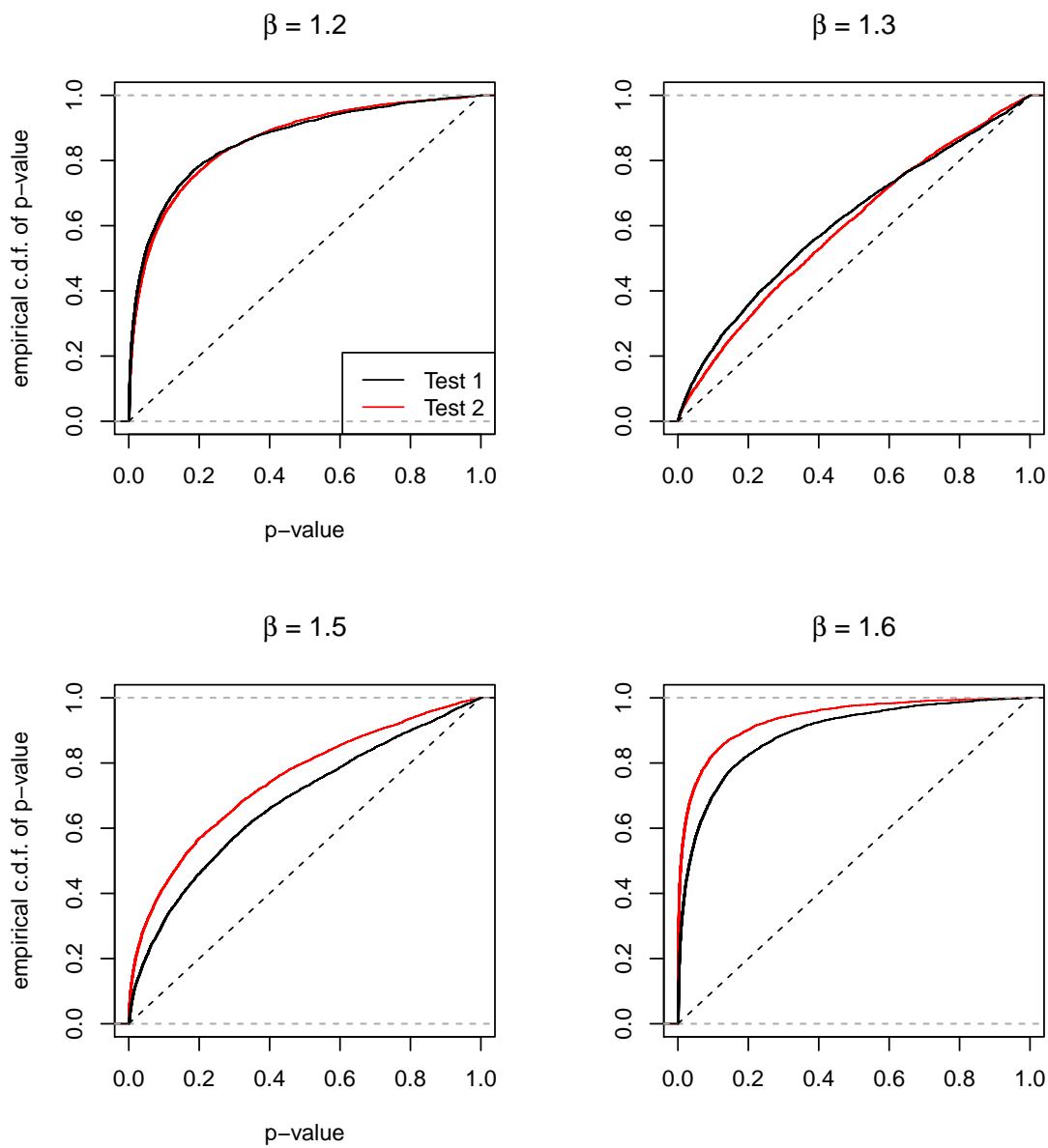
Figure 2: Empirical c.d.f. of p-values of T_1 and T_2 for testing $H_0 : \theta_0 = (2, 1.4)'$ under several alternatives of the form $\theta = (2, \beta)'$; 5000 replications with $N = 250$, $n = 817$.
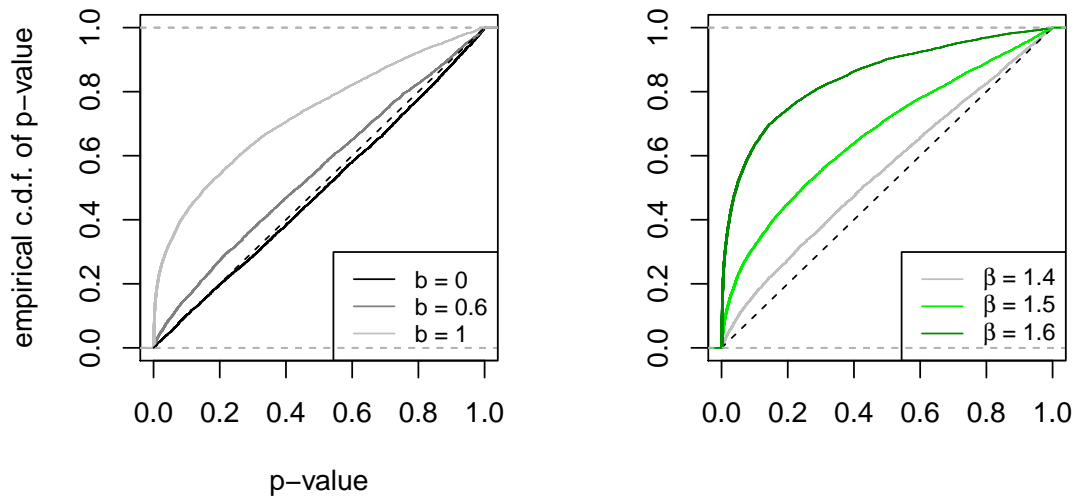
Figure 3: [left] Empirical c.d.f. of p-values of T_1 under $H_0 \colon \theta_0 = (2, 1.4)'$ for different dependence structure between AR(1) series : $b_i = b$ and $c_i = \sqrt{1 - b^2}$ and $N = 250$, $n = 817$. [right] Empirical c.d.f. of p-values of T_1 for testing $H_0 : \theta_0 = (2, 1.4)'$. AR(1) series are driven by common innovations, i.e. $b_i = 1$, $c_i = 0$, for $\theta = (2, \beta)'$; 5000 replications with $N = 250$, $n = 5500$.

References

Beran, J., Schützner, M. and Ghosh, S. (2010). From short to long memory: Aggregation and estimation. *Computational Statistics and Data Analysis*, **54,** 2432–2442.

Burkholder, D. L. (1973). Distribution function inequalities for martingales. *Annals of Probability*, **1,** 19–42.

Celov, D., Leipus, R. and Philippe, A. (2007). Time series aggregation, disaggregation and long memory. *Lithuanian Mathematical Journal*, **47,** 379–393.

Celov, D., Leipus, R. and Philippe, A. (2010). Asymptotic normality of the mixture density estimator in a disaggregation scheme. *Journal of Nonparametric Statistics*, **22,** 425–442.

Giraitis, L., Koul, H. L. and Surgailis, D. (2012). *Large Sample Inference for Long Memory Processes*. Imperial College Press, London.

Granger, C. W. J. (1980). Long memory relationship and the aggregation of dynamic models. *Journal of Econometrics*, **14,** 227–238.

Horváth, L. and Leipus, R. (2009). Effect of aggregation on estimators in AR(1) sequence. *TEST*, **18,** 546–567.

Jirak, M. (2013). Limit theorems for aggregated linear processes. *Advances in Applied Probability*, **45,** 520–544.

Leipus, R., Oppenheim, G., Philippe, A. and Viano, M.-C. (2006). Orthogonal series density estimation in a disaggregation scheme. *Journal of Statistical Planning and Inference*, **136,** 2547–2571.

Parzen, E. (1962). On estimation of probability density function and mode. *Annals of Mathematical Statistics*, **33,** 1065–1076.

Robinson, P. M. (1978). Statistical inference for a random coefficient autoregressive model. *Scandinavian Journal of Statistics*, **5,** 163–168.

Rosenthal, H. P. (1970). On the subspaces of L^p ($p > 2$) spanned by the sequences of independent random variables. *Israel Journal of Mathematics*, **8,** 273–303.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

von Bahr, B. and Esséen, C.-G. (1965). Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Annals of Mathematical Statistics*, **36,** 299–303.

6   Appendix: some proofs and auxiliary lemmas

We use the following martingale moment inequality.

Lemma 6.1 *Let $p > 1$ and $\{\xi_j, j \geq 1\}$ be a martingale difference sequence:* $\mathrm{E}[\xi_j | \xi_1, \ldots, \xi_{j-1}] = 0$, $j = 2, 3, \ldots$ *with* $\mathrm{E}|\xi_j|^p < \infty$. *Then there exists a constant $C_p < \infty$ depending only on p and such that*

$$\mathrm{E}\Big| \sum_{j=1}^{\infty} \xi_j \Big|^p \;\leq\; C_p \begin{cases} \sum_{j=1}^{\infty} \mathrm{E}|\xi_j|^p, & 1 < p \leq 2, \\ \big(\sum_{j=1}^{\infty} (\mathrm{E}|\xi_j|^p)^{2/p} \big)^{p/2}, & p > 2. \end{cases} \tag{6.1}$$

For $1 < p << \leq 2$, inequality (6.1) is known as von Bahr and Esséen inequality, see von Bahr and Esséen (1965), and for $p > 2$, it is a consequence of the Burkholder and Rosenthal inequality (Burkholder (1973), Rosenthal (1970), see also Giraitis *et al.* (2012, Lemma 2.5.2)).

PROOF OF PROPOSITION 2.1. Since $\widehat{a}_n$ in (2.5) is invariant w.r.t. a scale factor of innovations $\{\zeta(t)\}$, w.l.g. we can assume $b^2 + c^2 = 1$ and $\mathrm{E}\zeta^2(0) = 1$, $\mathrm{E}|\zeta(0)|^{2p} < \infty$. Then $\widehat{a}_n - a = \delta_{n1} + \delta_{n2}$, where

$$\delta_{n1} \;:=\; -\frac{aX^2(n)}{\sum_{t=1}^{n} X^2(t)}, \quad \delta_{n2} \;:=\; \frac{\sum_{t=1}^{n-1} X(t)\zeta(t+1)}{\sum_{t=1}^{n} X^2(t)}. \tag{6.2}$$

The statement of the proposition follows from

$$\mathrm{P}(|\delta_{ni}| > \gamma) \leq C(n^{-1} + n^{-(p/2)\wedge(p-1)}\gamma^{-p}) \qquad (0 < \gamma < 1, \ i = 1, 2). \tag{6.3}$$

To show (6.3) for $i = 1$, note that $\mathrm{P}(|\delta_{n1}| > \gamma) \leq \mathrm{P}(|S_{n1}| > n/2) + \mathrm{P}(S_{n2} > n\gamma/2)$, where $S_{n1} := \sum_{t=1}^{n}((1 - a^2)X^2(t) - 1)$, $S_{n2} := (1 - a^2)X^2(n)$. Thus, (6.3) for $i = 1$ follows from

$$\mathrm{E}|S_{n1}|^{p\wedge 2} \;\leq\; Cn \qquad \text{and} \qquad \mathrm{E}|S_{n2}|^p \;\leq\; C. \tag{6.4}$$

Consider the first relation in (6.4). Clearly, it suffices to prove it for $1 < p \leq 2$ only. We have $S_{n1} = 2S'_{n1} + S''_{n1}$, where

$$S'_{n1} \;:=\; (1 - a^2) \sum_{s_2 < s_1 \leq n} \sum_{t=1\vee s_1}^{n} a^{2(t-s_1)} a^{s_1 - s_2} \zeta(s_1)\zeta(s_2),$$

$$S''_{n1} \;:=\; (1 - a^2) \sum_{s \leq n} \sum_{t=1\vee s}^{n} a^{2(t-s)}(\zeta^2(s) - 1).$$

We will use the following elementary inequality: for any $-1 \leq a \leq 1, n \geq 1, s \leq n$

$$\alpha_n(s) \;:=\; (1 - a^2) \sum_{t=1\vee s}^{n} a^{2(t-s)} \;=\; \begin{cases} a^{2(1-s)}(1 - a^{2n}), & s \leq 0, \\ 1 - a^{2(n+1-s)}, & 1 \leq s \leq n \end{cases}$$

$$\leq\; C \begin{cases} a^{-2s}\min(1, 2n(1 - |a|)), & s \leq 0, \\ 1, & 1 \leq s \leq n. \end{cases} \tag{6.5}$$

Using the independence of $\{\zeta(s)\}$ and a and inequality (6.1) (twice) for $1 < p \leq 2$ we obtain

$$\begin{aligned} \mathrm{E}|S'_{n1}|^p &= \mathrm{E}\Big| \sum_{s_1 \leq n} \alpha_n(s_1)\zeta(s_1) \sum_{s_2 < s_1} a^{s_1 - s_2}\zeta(s_2) \Big|^p \\ &\leq CE \sum_{s_1 \leq n} \Big|\alpha_n(s_1)\zeta(s_1) \sum_{s_2 < s_1} a^{s_1 - s_2}\zeta(s_2)\Big|^p \\ &\leq CE \sum_{s_1 \leq n} |\alpha_n(s_1)|^p \sum_{s_2 < s_1} |a|^{p(s_1 - s_2)} \\ &\leq CE(1 - |a|)^{-1} \sum_{s \leq n} |\alpha_n(s)|^p \;\leq\; Cn \end{aligned}$$

15

since $\mathrm{E}(1 - |a|)^{-1} < \infty$ (see (2.3)) and $\sum_{s \leq n} |\alpha_n(s)|^p \leq Cn$ follows from (6.5). Similarly, since $\{\zeta^2(s) - 1,\ s \leq n\}$ form a martingale difference sequence,

$$\mathrm{E}|S''_{n1}|^p \quad \leq \quad C\mathrm{E}\sum_{s \leq n} |\alpha_n(s)|^p \quad \leq \quad Cn,$$

proving the first inequality (6.4).

Next, consider the second inequality in (6.4). We have $S_{n2} = 2S'_{n2} + S''_{n2}$, where

$$S'_{n2} \quad := \quad (1 - a^2) \sum_{s_2 < s_1 \leq n} a^{2(n-s_1)} a^{s_1 - s_2} \zeta(s_1)\zeta(s_2), \qquad S''_{n2} \quad := \quad (1 - a^2)\sum_{s \leq n} a^{2(n-s)}(\zeta^2(s) - 1).$$

Similarly as above, we obtain $\mathrm{E}|S''_{n2}|^p \leq C\mathrm{E}\sum_{s \leq n} |(1 - a^2)a^{-2(n-s)}|^p \leq C$ and $\mathrm{E}|S'_{n2}|^p \leq C\mathrm{E}\sum_{s_2 < s_1 \leq n} |(1 - a^2)|a|^{2(n-s_1)}|a|^{s_1 - s_2}|^p \leq C\mathrm{E}(1 - |a|)^{p-2} < \infty$, proving (6.4) and hence (6.3) for $i = 1$.

Consider (6.3) for $i = 2$. We have $\delta_{n2} = R_n/(n + S_{n1})$, where $R_n := (1 - a^2)\sum_{t=1}^{n-1} X(t)\zeta(t + 1)$ and $S_{n1} = \sum_{t=1}^{n}((1 - a^2)X^2(t) - 1)$ is the same as in (6.4). Then $\mathrm{P}(|\delta_{n2}| > \gamma) \leq \mathrm{P}(|R_n| > n\gamma/2) + \mathrm{P}(|S_{n1}| > n/2)$, where

$$\mathrm{P}(|S_{n1}| > n/2) \leq (n/2)^{-(p\wedge 2)}\mathrm{E}|S_{n1}|^{p\wedge 2} \leq C \begin{cases} n^{-(p-1)}, & 1 < p \leq 2, \\ n^{-1}, & p > 2, \end{cases}$$

according to (6.4). Therefore (6.3) for $i = 2$ follows from

$$\mathrm{E}|R_n|^p \leq C \begin{cases} n, & 1 < p \leq 2, \\ n^{p/2}, & p > 2. \end{cases} \tag{6.6}$$

Since $R_n = (1 - a^2)\sum_{s \leq n-1} \zeta(s)\sum_{t=1\vee s}^{n-1} a^{t-s}\zeta(t + 1)$ is a sum of martingale differences, by inequality (6.1) with $1 < p \leq 2$ we obtain

$$\begin{aligned}
\mathrm{E}|R_n|^p \quad &\leq \quad C\mathrm{E}\sum_{s \leq n-1} \Big|(1 - a^2)\zeta(s)\sum_{t=1\vee s}^{n-1} a^{t-s}\zeta(t + 1)\Big|^p \\
&\leq \quad C\mathrm{E}|1 - a^2|^p \sum_{s \leq n-1}\sum_{t=1\vee s}^{n-1} |a|^{p(t-s)} \\
&\leq \quad C\mathrm{E}|1 - a^2|^2 \Big(\sum_{s \leq 0} |a|^{-ps}\sum_{t=1}^{n-1} |a|^{pt} + \sum_{s=1}^{n-1}\sum_{t=s}^{n-1} |a|^{p(t-s)}\Big) \\
&\leq \quad C\mathrm{E}|1 - a^2|^2 \big\{(1 - |a|^p)^{-2} + n(1 - |a|^p)^{-1}\big\} \quad \leq \quad Cn,
\end{aligned}$$

proving (6.6) for $p \leq 2$. Similarly, using (6.1) with $p > 2$ we get

$$\begin{aligned}
\mathrm{E}|R_n|^p \quad &= \quad \mathrm{E}|1 - a^2|^p \mathrm{E}\Big[\big|\sum_{s \leq n-1} \zeta(s)\sum_{t=1\vee s}^{n-1} a^{t-s}\zeta(t + 1)\big|^p \big| a\Big] \\
&\leq \quad C\mathrm{E}|1 - a^2|^p \Big\{\sum_{s \leq n-1} \big(\mathrm{E}\big[\big|\zeta(s)\sum_{t=1\vee s}^{n-1} a^{t-s}\zeta(t + 1)\big|^p \big| a\big]\big)^{2/p}\Big\}^{p/2} \\
&\leq \quad C\mathrm{E}|1 - a^2|^p \Big\{\sum_{s \leq n-1}\sum_{t=1\vee s}^{n-1} a^{2(t-s)}\Big\}^{p/2} \\
&\leq \quad C\mathrm{E}|1 - a^2|^p \Big\{\sum_{s \leq 0} a^{-2s}\sum_{t=1}^{n-1} a^{2t} + \sum_{s=1}^{n-1}\sum_{t=s}^{n-1} a^{2(t-s)}\Big\}^{p/2} \\
&\leq \quad C\mathrm{E}|1 - a^2|^p \big\{(1 - a^2)^{-2} + n(1 - a^2)^{-1}\big\}^{p/2} \quad \leq \quad Cn^{p/2},
\end{aligned}$$

proving (6.6), (6.3) and the proposition. □

Let $a, a_1, \ldots, a_N$ be i.i.d. r.v.s with d.f. $G(x) = P(a \leq x)$ supported by $[-1, 1]$. Define $\widehat{G}_N(x) := N^{-1} \sum_{i=1}^{N} \mathbf{1}(a_i \leq x)$, $U_N(x) := N^{1/2}(\widehat{G}_N(x) - G(x))$, $x \in \mathbb{R}$, and $\omega_N(\delta)$ (= the modulus of continuity of U_N) by

$$\omega_N(\delta) := \sup_{0 \leq y - x \leq \delta} |U_N(y) - U_N(x)|, \quad \delta > 0.$$

Lemma 6.2 *Assume that G satisfies assumption A_6. Then for all $\epsilon > 0$,*

$$\epsilon^4 P(\omega_N(\delta) > 6\epsilon) \leq (3 + 3C)L_G \delta^\varrho + N^{-1},$$

where C is a constant independent of ϵ, δ, N.

PROOF. As in Shorack and Wellner (1986, p. 110, Inequality 1) we have that

$$E|U_N(y) - U_N(x)|^4 \leq 3P(a \in (x, y])^2 + N^{-1}P(a \in (x, y]),$$
$$E|U_N(y) - U_N(x)|^2|U_N(z) - U_N(y)|^2 \leq 3P(a \in (x, y])P(a \in (y, z]),$$

for $-1 \leq x \leq y \leq z \leq 1$. Now fix $\delta > 0$ and split $[-1, 1] = \cup_i \Delta_i$, where $\Delta_i = [-1 + i\delta, -1 + (i + 1)\delta]$, $i = 0, 1, \ldots, \lfloor 2/\delta \rfloor - 1$, $\Delta_{\lfloor 2/\delta \rfloor} = [-1 + \lfloor 2/\delta \rfloor \delta, 1]$. According to Shorack and Wellner (1986, p. 49, Lemma 1), for all $\epsilon > 0$,

$$\epsilon^4 P(\omega_N(\delta) > 6\epsilon) \leq (3 + 3C) \max_i P(a \in \Delta_i) + N^{-1},$$

where C is a constant independent of ϵ, δ, N. Lemma follows from Assumption A_6 on d.f. G of a r.v. a. □

Note that if we take $\delta = \delta_N = o(1)$, we then get $P(\omega_N(\delta) > \epsilon) \to 0$ as $N \to \infty$.

Lemma 6.3 *Let $\widehat{a}_{1,n}, \widehat{a}_{2,n}$ be given in (3.3) under assumptions A_1–A_6 with $\varrho = 1$. Then for all $\gamma \in (0, 1)$ and $n \geq 1$, it holds*

$$\sup_{x, y \in [-1, 1]} |P(\widehat{a}_{1,n} \leq x, \widehat{a}_{2,n} \leq y) - P(\widehat{a}_{1,n} \leq x)P(\widehat{a}_{2,n} \leq y)| = O(n^{-((p/2) \wedge (p-1))/(1+p)}).$$

PROOF. Define $\delta_{i,n} := \widehat{a}_{i,n} - a_i$, $i = 1, 2$. For $\gamma \in (0, 1)$, we have

$$P(|\delta_{1,n}| > \gamma \text{ or } |\delta_{2,n}| > \gamma) \leq P(|\delta_{1,n}| > \gamma) + P(|\delta_{2,n}| > \gamma) \leq C(n^{-(p/2) \wedge (p-1)}\gamma^{-p} + n^{-1}),$$

by Proposition 2.1. Consider now

$$\begin{aligned}
P(\widehat{a}_{1,n} \leq x, \widehat{a}_{2,n} \leq y) &= P(a_1 + \delta_{1,n} \leq x, a_2 + \delta_{2,n} \leq y) \\
&\leq P(a_1 + \delta_{1,n} \leq x, a_2 + \delta_{2,n} \leq y, |\delta_{1,n}| \leq \gamma, |\delta_{2,n}| \leq \gamma) + P(|\delta_{1,n}| > \gamma \text{ or } |\delta_{2,n}| > \gamma).
\end{aligned}$$

Then

$$\begin{aligned}
P(a_1 + \delta_{1,n} \leq x, a_2 + \delta_{2,n} \leq y, |\delta_{1,n}| \leq \gamma, |\delta_{2,n}| \leq \gamma) &\leq P(a_1 \leq x + \gamma, a_2 \leq y + \gamma, |\delta_{1,n}| \leq \gamma, |\delta_{2,n}| \leq \gamma) \\
&\leq G(x + \gamma)G(y + \gamma)
\end{aligned}$$

and

$$\begin{aligned}
P(a_1 + \delta_{1,n} \leq x, a_2 + \delta_{2,n} \leq y, |\delta_{1,n}| \leq \gamma, |\delta_{2,n}| \leq \gamma) &\geq P(a_1 \leq x - \gamma, a_2 \leq y - \gamma, |\delta_{1,n}| \leq \gamma, |\delta_{2,n}| \leq \gamma) \\
&\geq G(x - \gamma)G(y - \gamma) - P(|\delta_{1,n}| > \gamma \text{ or } |\delta_{2,n}| > \gamma).
\end{aligned}$$

17

From (2.7) we obtain

$$|G(x \pm \gamma)G(y \pm \gamma) - G(x)G(y)| \;=\; |(G(x) + O(\gamma))(G(y) + O(\gamma)) - G(x)G(y)| \;\leq\; C\gamma.$$

Hence

$$|\mathrm{P}(a_1 \leq x,\, a_2 \leq y) - G(x)G(y)| \;\leq\; C(\gamma + n^{-1} + n^{-(p/2)\wedge(p-1)}\gamma^{-p}). \tag{6.7}$$

In a similar way,

$$|\mathrm{P}(a_1 \leq x)\mathrm{P}(a_2 \leq y) - G(x)G(y)| \;\leq\; C(\gamma + n^{-1} + n^{-(p/2)\wedge(p-1)}\gamma^{-p}). \tag{6.8}$$

By (6.7), (6.8), the proof of the lemma is complete with $\gamma = \gamma_n = o(1)$, which satisfies $\gamma_n \sim n^{-(p/2)\wedge(p-1)}\gamma_n^{-p}$.
□